ОБОБЩЕННЫЙ ГРАДИЕНТНЫЙСПУСК С РАСТЯЖЕНИЕМ ПРОСТРАНСТВА. НАЧАЛО РАЗВИТИЯ

В.И.БИЛЕЦКИЙ
Институт кибернетики НАН Украины,
г. Киев, Украина,
(bilvassa@ukr.net)

На сегодняшний день алгоритмы с растяжением пространства стали широко известными как в нашей стране, так и за ее пределами. Особую популярность получили алгоритмы растяжения пространства в направлении разности двух последовательных алгоритмов при решении многих важных практических задач. А началом развития этих методов является метод растяжения пространства в направлении градиента (обобщенного градиента). В докладе пойдет речь об истоках развития методов растяжения пространства в задачах оптимизации выпуклых функций. Описывается схема алгоритма градиентного спуска с растяжением пространства в направлении градиента, а также приводятся некоторые примеры экспериментальных расчетов.

***Ключевые слова**: функция, метод, градиент, растяжение, спуск, алгоритм, точка минимума, коэффициент растяжения, шаг сдвига.*

Пусть рассматривается задача оптимизации выпуклой функции $f(x)$ в n – мерном евклидовым пространстве E_n. От функции $f(x)$ требуется всего лишь, чтобы она была непрерывной и выпуклой.

Идея растяжения пространства в задачах оптимизации выпуклых функций у Н.З.Шора возникла в 60-х годах прошлого столетия в связи с некоторыми проблемами нахождения точки минимума функции овражного типа и начала успешно развиваться во второй половине тех лет. Если функция $f(x)$ является сущест-

венно овражной, т.е. функцией с сильно вытянутыми линиями уровня вдоль русла оврага, то, например, обычный обобщенный градиентный спуск не гарантирует даже линейной скорости сходимости. Более того, может быть такое, что после какого-то количества шагов никакого улучшения ни по функционалу, ни по приближению к точке минимума не происходит. Идет процесс "топтания на месте", потому что направление спуска образует угол, близкий к $\pi/2$ с направлением на точку минимума.

Чтобы устранить этот недостаток и добиться в задачах оптимизации функций овражного типа результата, т.е. найти точку минимума, Н.З.Шор придумал способ ускорения сходимости и нашел метод гарантированного нахождения точки минимума за приемлемое время. Таким способом оказалась процедура преобразования пространства E_n таким образом, чтобы угол между антиградиентом и направлением на точку минимума уменьшился. А этого можно было достичь путем растяжения пространства E_n в направлении градиента функции $f(x)$ в рассматриваемой точке.

Так возник и теоретически был обоснован метод обобщенного градиентного спуска с растяжением пространства в направлении градиента (ОГСРП) [1]. В работе [1] вводится понятие оператора растяжения пространства в некотором направлении ξ с некоторым заданным коэффициентом растяжения α, который представляет собой матрицу преобразования пространства аргументов, элементы которой определяются специальным образом от значений величин ξ и α. Приводится также схема алгоритма ОГСРП, суть которой сводится к следующему.

0-й шаг. Пусть задано начальное приближение $x_0 \in E_n$ и единичная матрица $B_0^* = E$.

На k-ом шаге ($k = 1, 2, \ldots$) последовательно вычисляем:

1) $g_f(x_k)$ - обобщенный градиент функции $f(x)$ в точке x_k;

2) $\eta_k = B_k^* g_f(x_k)$ - градиент в растянутом пространстве;

3) $\xi_k = \dfrac{\eta_k}{\|\eta_k\|}$;

4) $x_{k+1} = x_k - h_k B_k \xi_k$;

5) $B_{k+1} = B_k R_{\beta_k}(\xi_k)$.

В алгоритме, описанном в п.1-5 $R_{\beta_k}(\xi_k)$ - оператор растяжения пространства в направлении $\xi = \xi_k$ с коэффициентом $\beta_k = 1/\alpha_k$, α_k - коэффициент растяжения пространства в направлении градиента $g_f(x_k)$ в точке x_k, h_k - шаг сдвига.

В работе [1] доказано, что при применении алгоритма, описанного в п. 1-5, последовательность $\{x_k\}$, $k = 1, 2, \ldots$, сходится к точке минимума x^* и $\lim\limits_{k\to\infty} f(x_k) = f(x^*)$.

В п. 1-5 описан класс алгоритмов, отличающиеся друг от друга различными теоретически обоснованными способами задания коэффициента растяжения α_k и шага сдвига $h_k, k = 1, 2, \ldots$, обеспечивающие сходимость метода.

В подтверждение теоретического обоснования метода ОГСРП и его эффективности были проведены численные эксперименты на многих примерах, которые показали, что разработанный метод является эффективным даже в тех случаях, для которых другие методы либо сталкиваются в процессе оптимизации с большими проблемами, либо не дают положительного результата вообще.

Ниже приводятся примеры численных экспериментов для различных функций, в каждом из которых указано начальное приближение x^0, способ задания значений h_k и α_k, количество итераций, за которое был получен приемлемый результат.

1. $f(x) = x_1^2 + |x_2|^3$. Точка минимума - $x^* = (0, 0)$.

$x_1^0 = 1,\ x_2^0 = 1,\ h_k = h = 1,\ \alpha_k = \alpha = 5/3$.

После 40-а итераций получен результат:

$x_1 = 0.0000044,\ x_2 = 0.000127$.

2. $f(x)=\max\left\{(x_1^2+x_2^2),10\left[(x_1-1)^2+x_2^2\right]\right\}$. $x^*=(0.7598,\ 0)$.

$x_1^0=0,\ x_2^0=1,\ h_k=h=1,\ \alpha_k=\alpha=2.$

После 23-х итераций получен результат:

$x_1=0.75986,\ \ x_2=\ 0.0051.$

3. $f(x)=x_1^2+10x_2^2+30x_3^2+50x_4^2+90x_5^2+100x_6^2$.

Точка минимума $x^*=(0,0,0,0,0,0)$.

$x^0=(1,\ 1,\ 1,\ 1,\ 1,\ 1).\ \ h_k=h=1,\ \alpha_k=\alpha=2.$

После 100-а итераций получен результат:

$x_1=-1.42\cdot10^{-5},\ \ x_2=-2.48\cdot10^{-6},\ \ x_3=3.23\cdot10^{-7},$

$x_4=2.34\cdot10^{-6},\ \ x_5=-5.8\cdot10^{-8},\ \ \ x_6=1.08\cdot10^{-6}.$

4. Функция из примера 3 оптимизировалась при тех же условиях за исключением шага сдвига, который на каждом шаге вычислялся по формуле $h_k=\dfrac{2f(x_k)}{\|\eta_k\|}$.

После 50-и итераций получен результат:

$x_1=1.84\cdot10^{-9},\ \ \ x_2=3.21\cdot10^{-10},\ \ x_3=1.6\cdot10^{-10},$

$x_4=3.81\cdot10^{-10},\ \ x_5=8.52\cdot10^{-11},\ \ \ x_6=7.87\cdot10^{-11}.$

5. Решить систему 3-х уравнений с 3-я неизвестными:

$$\begin{cases} S_1=x_1+x_2+x_3-1=0, \\ S_2=0.5x_1^2+0.2x_2^3-0.5x_3+0.5=0, \\ S_3=x_1+x_2+0.5x_3^2-0.5=0. \end{cases}$$

Решением этой системы являются значения неизвестных

$x_1=0,\ x_2=0,\ x_3=1.$

С помощью метода ОГСРП оно было найдено как результат решения эквивалентной задачи, а именно, задачи минимизации функции $\varphi=\max\limits_i S_i^2$, $i=1,2,3$. Начальным приближением были значения неизвестных $x_1^0=0.2,\ x_2^0=0.2,\ x_3=0.5$. Задача реша-

лась с постоянными $h_k = h = 3.5,\ \alpha_k = \alpha = 2.$ За 42 итерации получено решение:

$x_1 = 3.45 \cdot 10^{-6},\quad x_2 = -8.62 \cdot 10^{-6},\ x_3 = 1.000005.$

Необходимо отметить, что при известном значении функции $f(x^*)$ в точке минимума более эффективным по скорости сходимости является (и это показали многочисленные экспериментальные расчеты на тестовых примерах) алгоритм ОГСРП с автоматической регулировкой на каждой итерации шага $h_k, k = 1, 2, \ldots$, т.е. вычислением его по формуле

$$h_k = \frac{\gamma\left[f(x_k) - f(x^*)\right]}{\|\eta\|}, \qquad (1)$$

где γ - некоторый (заданный до начала процесса) коэффициент, $f(x^*)$ - значение функции в точке минимума, $f(x_k)$ - значение функции в текущей точке x_k, η - градиент в растянутом пространстве в точке x_k, чем задание его как $h_k = h = \text{const}$.

Но не во всех случаях значение функции в точке минимума известно. Тогда формулу (1) для вычисления шага на каждой итерации применить нельзя. Но выход есть. И для ускорения сходимости при неизвестном значении функции в точке минимума для минимизации можно применить алгоритм поиска с заданным шагом $h_k = h = \text{const}$, либо применить описанный выше алгоритм поиска точки минимума с переменным шагом, вычисляемого по формуле [2]

$$h_k = \frac{\gamma\left[f(x_k) - m_k\right]}{\|\eta\|}, \qquad (2)$$

где значения m_k, $k = 1, 2, \ldots,$ вычисляются определенным способом и такие, что $m_k \xrightarrow[k\to\infty]{} m^* = f(x^*)$. В [3] описана модификация алгоритма ОГСРП с неизвестны значением m^* и выбором

шага по формуле (2), которая была апробирована на приведенных ниже примерах и показала более высокую эффективность.

1. $F_1(X)=\left(\frac{x_1-3}{100}\right)-(x_2-x_1)+\exp 20(x_2-x_1)$,

2.

$$F_2(X)=(x_1+10x_2)^2+5(x_3-x_4)^2+(x_2-2x_3)^4+10(x_1-x_4)^4.$$

К сожалению, алгоритмы на основе метода растяжения пространства в направлении градиента не получили широкого применения для решения реальных практических задач, но метод обобщенного градиентного спуска с растяжением пространства в направлении градиента стал отправной точкой и дал толчок на развитие алгоритмов с растяжением пространства в направлении разности двух последовательных градиентов [4], которые уже к концу 70-х годов прошлого столетия получили широкое применение для решения ряда практических задач большого объема [5].

Литература

1. Шор Н.З., Билецкий В.И. Метод растяжения пространства для ускорения сходимости в задачах овражного типа // Теория оптимальных решений. – Киев: Ин-т кибернетики АН УССР, - 1969. – вып. 2. – С. 3-18.
2. Шор Н.З. О скорости сходимости метода обобщенного градиентного спуска с растяжением пространства // Кибернетика, - 1970. - № 2. – С. 80-85.
3. Билецкий В.И. К методу обобщенного градиентного спуска с растяжением пространства // Теория оптимальных решений. – Киев: Ин-т кибернетики АН УССР, - 1972. - С. 20-25.
4. Шор Н.З., Журбенко Н.Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов // Кибернетика, - 19170. - № 3. – С. 51-59.
5. Шор Н.З. Методы минимизации недифференцируемых функций и их приложения – К.: Наукова думка, -1979.-199 с.